Meaningful Representation and Recombination of Variable Length Genomes

Matt Ryerkerk, Ron Averill, Kalyanmoy Deb Department of Mechanical Engineering Michigan State University East Lansing, Michigan

{ryerkerk, averillr, kdeb}@egr.msu.edu

ABSTRACT

Optimization algorithms typically operate only within a fixedsized design space, solving problems with a fixed number of parameters. However, many optimization problems allow for a variable number of components, where the optimal number may not be known a priori. These problems may be solved by using a genetic algorithm that utilizes a variable-length genome. A particular challenge when using variable-length genomes is the recombination of two parent solutions to produce meaningful children. The performances of several crossover operators are investigated and compared using a sensor placement testbed problem. It is shown that the optimal number of sensors may be determined by each operator, and that performance is improved when care is taken to preserve similarities between parent solutions. Performance may also be further improved by introducing a bias when pairing parents for recombination based on their relative genome lengths.

Categories and Subject Descriptors

I.2.8 [Artificial Intelligence]: Problem Solving, Control Methods and Search

General Terms

Algorithms

Keywords

Genetic Algorithms, Representations, Recombination Operators

1. INTRODUCTION

A key to the success of a genetic algorithm is its ability to form building blocks, short subsequences (or schemata) of the genome that have a positive influence on the objective value of an individual. For a population to converge toward optimal solutions, these building blocks must be able to propagate through successive generations. The genetic algorithm operators must be chosen such that they do not hinder this process.

Genetic algorithms typically solve problems with a fixed number of parameters. This paper considers problems that contain a variable number of components, in particular a sensor placement problem. Sensors are to be placed such that the coverage is maximized while minimizing the cost. In this type of problem it is unlikely that the optimal number of sensors is known *a priori*. A

Copyright is held by the author/owner(s). *GECCO'12 Companion*, July 7–11, 2012, Philadelphia, PA, USA. ACM 978-1-4503-1178-6/12/07.

Erik Goodman BEACON Center for the Study of Evolution in Action Michigan State University East Lansing, Michigan

goodman@egr.msu.edu

variable length genome can be used to optimize not only the placement of the sensors but the number of sensors used.

The idea of building blocks remains; the optimal placement of a particular sensor is dependent on the positioning of other nearby sensors. Preserving these relationships increases the likelihood of producing meaningful children during recombination. Standard crossover operators, such as an n-point crossover, are inefficient for this type of problem as they make no effort to preserve these relationships while exchanging information.

A survey on wireless sensor network deployment includes several genetic algorithms [3]. These studies typically either fix the number of sensors, allow the number of sensors to change through mutation, and/or modify an n-point crossover to work with genomes of differing lengths. This paper investigates several recombination operators that allow the number of sensors in a solution to vary.

2. PROBLEM FOMULATION

This paper uses a sensor placement problem as a testbed for the proposed recombination operators. Sensors are to be placed on a square field such that the coverage is maximized while minimizing the cost. The strength, or effective sensing radius, of each individual sensor is also optimized. The cost of a sensor is a combination of a fixed base cost plus a cost that increases with the sensor's effective range. An objective function is defined which combines the cost of the sensor network and a penalty that scales linearly with the portion of the field that is left uncovered. A point is said to be uncovered if it is not within the sensing range of at least one sensor. This penalty term is weighted heavily such that the optimal solutions will achieve complete, or near complete, coverage of the field. A sample solution is shown in Figure 1.

3. PROPOSED METHODOLOGIES

A multi-objective genetic algorithm is employed to minimize the objective function defined in Section 2 while also minimizing the number of sensors used.

Three different recombination operators are demonstrated and compared. Each operator is demonstrated twice, using one of two parental pairing techniques. The first type randomly pairs parents together for recombination. The second type adaptively pairs parents together based on the difference in their genome lengths. This is done such that the majority of pairs have relative genome lengths that were found to be most successful in previous generations.



Figure 1. Sample solution with an objective function value of 48.85 using 28 sensors.

The genome used in each methodology contains a variable number of genes. Each gene contains the real-valued parameters describing a single component, in this case a sensor. Crossover can only occur between genes, not within a gene. The number of genes is not fixed and the number of genes in each child may not match the number of genes in either parent.

The algorithm is run 100 times for each combination of crossover operator and parent pairing technique. Each trial is run for 3000 generations using a population of 50.

3.1 Fixed-Length Representation

A fixed-length genome may be used to represent a problem with a variable number of components. This is done by allowing the sensing range of a sensor to be negative. A sensor with a negative range is not considered when evaluating the design. This method creates an upper bound on the total number of sensors in a single design. An n-point crossover is used for recombination.

3.2 Spatial Crossover

The performance of a sensor network is highly dependent on the location of sensors relative to each other. A spatial crossover is well suited for this type of problem [1]. The sensor field is split into two parts by a randomly oriented and positioned line. This line splits the sensors of each parent into two groups. Two new children are produced by inheriting one group from each parent.

3.3 Similar Gene Crossover

A crossover proposed by Hutt and Warwick [2], the Synapsing Variable-Length Crossover (SLVC), preserves common subsequences in the genomes of both parents. SLVC maintains the order of genes in the genome, but in this problem the order of sensors defined in the genome is inconsequential. The Similar Gene crossover is similar to SVLC, but the parent genomes are rearranged based on the similarity of gene pairs between the two parents. The similarity between two genes is found by comparing the parameter values they contain. The first genes in the rearranged parent genomes will be the most similar pair of genes between the two parents. The subsequent genes will be arranged in decreasing gene pair similarity. An n-point crossover is then applied.

4. RESULTS AND DISCUSSION

The average best objective function value at each generation for each algorithm is shown in Figure 2. The fixed-length



Figure 2. Best objective function values at each generation, averaged over 100 trials. Solid lines are trials using adaptive parent pairing (A), dashed lines use random pairing (R).

representation method is relatively slow to converge when compared to the other operators. This is a result of the n-point crossover. During this crossover genes remain in their original position in the genome, the first gene in a child's genome is also the first gene in one of the parent genomes. If the genes at each position in both parents are unrelated then the gene that the child inherits is effectively random and meaningless. Eventually the population will converge such that the gene at each position of the genome is similar among all individuals. The first gene in each individual may describe a sensor located in the upper-left corner, for example. This allows for meaningful crossovers.

The spatial crossover produces meaningful results by preserving the positional relationships among groups of sensors. The similar gene crossover ensures that similarities in the two parents are preserved in the children. These considerations help to preserve building blocks during recombination, and increase the rate of meaningful children produced. While the spatial crossover has the best performance it is important to note it may not always be apparent, or possible, to adapt this crossover for use in other problems.

All methods tested are shown to be viable and produce high quality solutions. The performance of each recombination operator is increased by the adaptive parent pairing. For this problem, the most successful parent pairings tend to be between parents with the same number of genes. The spatial and similar gene crossovers offer the best performance by allowing building blocks to be more easily preserved during recombination.

5. REFERENCES

- [1] Cherba, D., Punch, W. 2006. Crossover Gene Selection by Spatial Location. In *Proceedings of 8th Annual Conference* on Genetic and Evolutionary Computation (Seattle, WA, USA, Jul. 8-12, 2006). GECCO '06, ACM, New York, NY, 1111-1116. DOI= http://doi.acm.org/10.1145/161468.16147.
- [2] Hutt, B., and Warwick, K. 2007. Synapsing variable-length crossover: Meaningful crossover for variable-length genomes. *IEEE Transactions on Evolutionary Computation* 11, 1 (Feb. 2007), 118-131, 2007. DOI=http://dx.doi.org/10.1109/TEVC.2006.878096
- [3] Marks, M. 2010. A Survey of Multi-Objective Deployment in Wireless Sensor Networks. *Journal of Telecommunications and Information Technology* (Mar. 2010), 36-41.